#19

09858332

1600

RAW SEQUENCE LISTING DATE: 12/27/2002
PATENT APPLICATION: US/09/858,332A TIME: 14:16:41

Input Set : D:\seqlist.txt
Output Set: N:\CRF4\12272002\I858332A.raw

<110> APPLICANT: Tchaga, Grigory S.
Jokhadze, George
<120> TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for Using the Same
<130> FILE REFERENCE: CLON-056CIP
<140> CURRENT APPLICATION NUMBER: US 09/858,332A
<141> CURRENT FILING DATE: 2001-05-15
<150> PRIOR APPLICATION NUMBER: 09/404,017
<151> PRIOR FILING DATE: 1999-09-23
<150> PRIOR APPLICATION NUMBER: 60/101,867
<151> PRIOR FILING DATE: 1998-09-25
<160> NUMBER OF SEQ ID NOS: 20
<170> SOFTWARE: FastSEQ for Windows Version 4.0
<210> SEQ ID NO: 1
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: affinity peptide
<400> SEQUENCE: 1
His Leu Ile His Asn Val His Lys Glu Glu His Ala His Ala His Asn
1 5 10 15
<210> SEQ ID NO: 2
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: affinity peptide
<400> SEQUENCE: 2
His Asp Asp His Asp Asp His Asp Asp His Asp Asp His Asp Asp His
1 5 10 15
Asp Asp
<210> SEQ ID NO: 3
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: affinity peptide
<400> SEQUENCE: 3
His Glu Glu His Glu Glu His Glu Glu His Glu Glu His Glu Glu His
1 5 10 15
<210> SEQ ID NO: 4
<211> LENGTH: 18

RAW SEQUENCE LISTING DATE: 12/27/2002
PATENT APPLICATION: US/09/858,332A TIME: 14:16:41

Input Set : D:\seqlist.txt
Output Set: N:\CRF4\12272002\I858332A.raw

<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: affinity peptide
<400> SEQUENCE: 4
His Asp Glu His Asp Glu His Asp Glu His Asp Glu His Asp Glu His
1 5 10 15
Asp Glu
<210> SEQ ID NO: 5
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: affinity peptide
<400> SEQUENCE: 5
His Glu Asp His Glu Asp His Glu Asp His Glu Asp His Glu Asp His
1 5 10 15
Glu Asp
<210> SEQ ID NO: 6
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: enterokinase cleavage site
<400> SEQUENCE: 6
Ile Glu Gly Arg
1
<210> SEQ ID NO: 7
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: a factor Xa cleavage site
<400> SEQUENCE: 7
Leu Val Pro Arg Gly Ser
1 5
<210> SEQ ID NO: 8
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: a thrombin cleavage site
<400> SEQUENCE: 8
His Pro Phe His Leu Val Ile His
1 5
<210> SEQ ID NO: 9
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence

RAW SEQUENCE LISTING DATE: 12/27/2002
PATENT APPLICATION: **US/09/858,332A** TIME: 14:16:41

Input Set : **D:\seqlist.txt**
Output Set: **N:\CRF4\12272002\I858332A.raw**

```
141 <220> FEATURE:
142 <223> OTHER INFORMATION: a renin cleavage site
144 <400> SEQUENCE: 9
145 Cys Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
146  1               5                  10
149 <210> SEQ ID NO: 10
150 <211> LENGTH: 8
151 <212> TYPE: PRT
152 <213> ORGANISM: Artificial Sequence
154 <220> FEATURE:
155 <223> OTHER INFORMATION: an immunological tag
157 <400> SEQUENCE: 10
158 Asp Tyr Lys Asp Asp Asp Asp Lys
159  1               5
162 <210> SEQ ID NO: 11
163 <211> LENGTH: 11
164 <212> TYPE: PRT
165 <213> ORGANISM: Artificial Sequence
167 <220> FEATURE:
168 <223> OTHER INFORMATION: an immunological tag
170 <400> SEQUENCE: 11
171 Cys Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
172  1               5                  10
175 <210> SEQ ID NO: 12
176 <211> LENGTH: 11
177 <212> TYPE: PRT
178 <213> ORGANISM: Artificial Sequence
180 <220> FEATURE:
181 <223> OTHER INFORMATION: an immunological tag
183 <400> SEQUENCE: 12
184 Cys Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu
185  1               5                  10
188 <210> SEQ ID NO: 13
189 <211> LENGTH: 3430
190 <212> TYPE: DNA
191 <213> ORGANISM: Artificial Sequence
193 <220> FEATURE:
194 <223> OTHER INFORMATION: DNA sequence of vector containing cDNA of
195       recombinant enterokinase
197 <400> SEQUENCE: 13
198 gacgaaaggg cctcgtgata cgcctatttt tataggttaa tgtcatgata ataatggttt 60
199 cttagacgtc aggtggcact tttcggggaa atgtgcgcgg aacccctatt tgtttatttt 120
200 tctaaataca ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat 180
201 aatattgaaa aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt 240
202 ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg 300
203 ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga 360
204 tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc 420
205 tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga gcaactcggt cgccgcatac 480
206 actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg 540
```

RAW SEQUENCE LISTING DATE: 12/27/2002
PATENT APPLICATION: US/09/858,332A TIME: 14:16:41

Input Set : D:\seqlist.txt
Output Set: N:\CRF4\12272002\I858332A.raw

207 gcatgacagt aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca 600
208 acttacttct gacaacgatc ggaggaccga aggagctaac cgctttttg cacaacatgg 660
209 gggatcatgt aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg 720
210 acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg 780
211 gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggataaag 840
212 ttgcaggacc acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg 900
213 gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct 960
214 cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac 1020
215 agatcgctga gataggtgcc tcactgatta agcattggta actgtcagac caagtttact 1080
216 catatatact ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga 1140
217 tccttttga taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt 1200
218 cagaccccgt agaaaagatc aaaggatctt cttgagatcc ttttttctg cgcgtaatct 1260
219 gctgcttgca aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc 1320
220 taccaactct ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc 1380
221 ttctagtgta gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc 1440
222 tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg 1500
223 ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggtt 1560
224 cgtgcacaca gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg 1620
225 agctatgaga aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg 1680
226 gbgcagggtc ggaacaggag agcgcacgag ggagcttcca gggggaaacg cctggtatct 1740
227 ttatagtcct gtcgggtttc gccacctctg acttgagcgt cgatttttgt gatgctcgtc 1800
228 agggggcgg agcctatgga aaaacgccag caacgcggcc tttttacggt tcctggcctt 1860
229 ttgctggcct tttgctcaca tgttctttcc tgcgttatcc cctgattctg tggataaccg 1920
230 tattaccgcc tttgagtgag ctgataccgc tcgccgcagc cgaacgaccg agcgcagcga 1980
231 gtcagtgagc gaggaagcgg aagagcgccc aatacgcaaa ccgcctctcc ccgcgcgttg 2040
232 gccgattcat taatgcagct ggcacgacag gtttcccgac tggaaagcgg gcagtgagcg 2100
233 caacgcaatt aatgtgagtt agctcactca ttaggcaccc caggctttac actttatgct 2160
234 tccggctcgt atgttgtgtg gaattgtgag cggataacaa tttcacacag gaaacagcta 2220
235 tgaccatgat tacgccaagc ttgaaggatc atctcatcca caatgtccac aaagaggagc 2280
236 acgctcatgc ccacaacaag atcgatattg tcggaggaag tgactccaga gaaggagcct 2340
237 ggccttgggt cgttgctctg tatttcgacg atcaacaggt ctgcggagct tctctggtga 2400
238 gcagggattg gctggtgtcg gccgcccact gcgtgtacgg gagaaatatg gagccgtcta 2460
239 agtggaaagc agtgctaggc ctgcatatgg catcaaatct gacttctcct cagatagaaa 2520
240 ctaggttgat tgaccaaatt gtcataaacc cacactacaa taaacggaga aagaacaatg 2580
241 acattgccat gatgcatctt gaaatgaaag tgaactacac agattatata cagcctattt 2640
242 gtttaccaga agaaaatcaa gtttttcccc caggaagaat ttgttctatt gctggctggg 2700
243 gggcacttat atatcaaggt tctactgcag acgtactgca agaagctgac gttccccttc 2760
244 tatcaaatga gaaatgtcaa caacagatgc cagaatataa cattacggaa aatatggtgt 2820
245 gtgcaggcta tgaagcagga ggggtagatt cttgtcaggg ggattcaggc ggaccactca 2880
246 tgtgccaaga aaacaacaga tggctcctgg ctggcgtgac gtcatttgga tatcaatgtg 2940
247 cactgcctaa tcgcccaggg gtgtatgccc gggtcccaag gttcacagag tggatacaaa 3000
248 gttttctaca tgagctcgta attagctgag aattcactgg ccgtcgtttt acaacgtcgt 3060
249 gagcctggga aaaccctggc gttacccaac ttaatcgcct tgcagcacat ccccctttcg 3120
250 ccagctggcg taatagcgaa gaggcccgca ccgatcgccc ttcccaacag ttgcgcagcc 3180
251 tgaatggcga atggcgcctg atgcggtatt ttctccttac gcatctgtgc ggtatttcac 3240
252 accgcatatg gtgcactctc agtacaatct gctctgatgc cgcatagtta agccagcccc 3300
253 gacacccgcc aacacccgct gacgcgccct gacgggcttg tctgctcccg gcatccgctt 3360
254 acagacaagc tgtgaccgtc tccgggagct gcatgtgtca gaggttttca ccgtcatcac 3420
255 cgaaacgcgc 3430

RAW SEQUENCE LISTING DATE: 12/27/2002
PATENT APPLICATION: **US/09/858,332A** TIME: 14:16:41

Input Set : **D:\seqlist.txt**
Output Set: **N:\CRF4\12272002\I858332A.raw**

```
257 <210> SEQ ID NO: 14
258 <211> LENGTH: 279
259 <212> TYPE: PRT
260 <213> ORGANISM: Artificial Sequence
262 <220> FEATURE:
263 <223> OTHER INFORMATION: protein sequence of vector containing cDNA of
264       recombinant enterokinase
266 <400> SEQUENCE: 14
267 Met Thr Met Ile Thr Pro Ser Leu Lys Asp His Leu Ile His Asn Val
268  1               5                  10                  15
269 His Lys Glu Glu His Ala His Ala His Asn Lys Ile Asp Ile Val Gly
270             20                  25                  30
271 Gly Ser Asp Ser Arg Glu Gly Ala Trp Pro Trp Val Val Ala Leu Tyr
272         35                  40                  45
273 Phe Asp Asp Gln Gln Val Cys Gly Ala Ser Leu Val Ser Arg Asp Trp
274     50                  55                  60
275 Leu Val Ser Ala Ala His Cys Val Tyr Gly Arg Asn Met Glu Pro Ser
276 65                  70                  75                  80
277 Lys Trp Lys Ala Val Leu Gly Leu His Met Ala Ser Asn Leu Thr Ser
278                 85                  90                  95
279 Pro Gln Ile Glu Thr Arg Leu Ile Asp Gln Ile Val Ile Asn Pro His
280             100                 105                 110
281 Tyr Asn Lys Arg Arg Lys Asn Asn Asp Ile Ala Met Met His Leu Glu
282         115                 120                 125
283 Met Lys Val Asn Tyr Thr Asp Tyr Ile Gln Pro Ile Cys Leu Pro Glu
284     130                 135                 140
285 Glu Asn Gln Val Phe Pro Pro Gly Arg Ile Cys Ser Ile Ala Gly Trp
286 145                 150                 155                 160
287 Gly Ala Leu Ile Tyr Gln Gly Ser Thr Ala Asp Val Leu Gln Glu Ala
288                 165                 170                 175
289 Asp Val Pro Leu Leu Ser Asn Glu Lys Cys Gln Gln Gln Met Pro Glu
290             180                 185                 190
291 Tyr Asn Ile Thr Glu Asn Met Val Cys Ala Gly Tyr Glu Ala Gly Gly
292         195                 200                 205
293 Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Cys Gln Glu
294     210                 215                 220
295 Asn Asn Arg Trp Leu Leu Ala Gly Val Thr Ser Phe Gly Tyr Gln Cys
296 225                 230                 235                 240
297 Ala Leu Pro Asn Arg Pro Gly Val Tyr Ala Arg Val Pro Arg Phe Thr
298                 245                 250                 255
299 Glu Trp Ile Gln Ser Phe Leu His Glu Leu Val Ile Ser Glu Phe Thr
300             260                 265                 270
301 Gly Arg Arg Phe Thr Thr Ser
302         275
305 <210> SEQ ID NO: 15
306 <211> LENGTH: 12
307 <212> TYPE: PRT
308 <213> ORGANISM: Artificial Sequence
310 <220> FEATURE:
```

VERIFICATION SUMMARY DATE: 12/27/2002
PATENT APPLICATION: **US/09/858,332A** TIME: 14:16:42

Input Set : **D:\seqlist.txt**
Output Set: **N:\CRF4\12272002\I858332A.raw**